## Parsed with column specification:  
## cols(  
## `States with Emission Caps` = col\_character(),  
## X2 = col\_character(),  
## X3 = col\_character(),  
## X4 = col\_character(),  
## X5 = col\_character(),  
## X6 = col\_character(),  
## X7 = col\_character()  
## )

## Goal:

We are a state legislation group trying to decide what would be the most effective use of our funds in order to decrease the CO2 emissions in our state. To do so, we're measuring the impact that various legislation and investments have had on CO2 emissions across states in the past in order to predict the most effective methods for future investment.

# Data and Measurements:

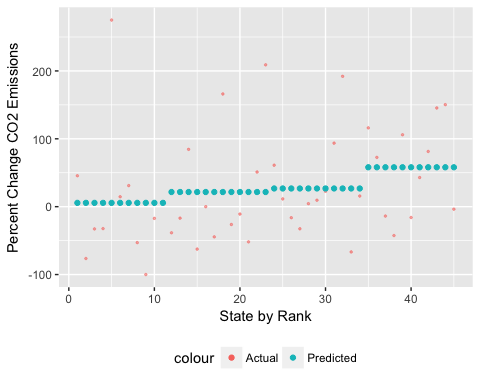
The data in this report is gleaned from the US Energy Information Administration, the US Environmental Protection Agency, the Bureau of Transportation Statistics, the US Department of Agriculture, and the National Conference of State Legislation. The binary variable of emissions cap law was coded based on the presence or absence of particular legislation enacted by state, as recorded in the National Conference of State Legislation. The investment in renewable energy is a dollar value representing total investment from 2002-2017; it represents total invesment instead of annual investment because many technological advances can be accomplished through large one-time investments instead of annual investments. In other words, using total investment is meant to account for technological improvements that may have happened in the past and reduce bias for certain states that may have had an abnormally large investment in one particular year. On the other hand, the public transportation variable is the amount of unlinked rides on public transportation by state in the year 2013. This is an annual variable because transportation is more likely to be consistent year to year. Additionally, Hawaii, Rhode Island, and Connecticut were dropped as variables because of insufficient data regarding CO2 emissions; the EPA does not currently report percent change emissions of these three states. There are two datasets for this project; one contains all states besides the three listed above, and the other omits Louisiana and Arkansas. This is because the percent change of LA and AR are substantially higher than the other states, with Lousiana having a 11200% increase in CO2 emissions from 1980-2014 and Arkansas having a 723% increase in CO2 emissions during that same time. Louisiana increased emissions by such a substantive amount in the past three decades because of an increased focus on oil and gas in the economy. Arkansas' increase is due in part to the success of Walmart, which is headquartered in the state. However, these two states do not stand out in terms of total CO2 emissions in any given year; rather, it is the percent change variable that makes them outliers becuase of their comparatively low emissions in the 1980s compared to their current emission rates.

# Univariate Graphic

Dependent Variable: CO2 Emissions by State Graphic: Rank of State Percent Change in CO2 Emissions from 1980-2014

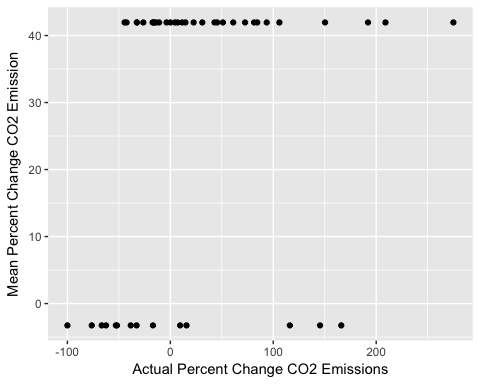
In the above figure, the interactive feature lists state rank of change in CO2 emissions, actual percent change in emissions, and state name. As the figure shows, the percent change in CO2 emissions from 1980-2014 ranges from a 100% decrease to 275% increase in emissions. Louisiana and Arkansas were omitted from this graph because they were outlier, skewing the visibility. LA had an increase in 11200% emissions, while AR had an increase of 723%. The mean change in CO2 emission by state during this time was 280.39% increase, and the median change in CO2 emissions was 9.52% increase. When the outliers of LA and AR are excluded, the mean change is 27.88% increase in CO2 emissions, and the median change is 7.05% increase.

# How does level of investment in renewable energy affect CO2 emission by state?



The variable for renewable energy was taken from the USDA and shows the total amount in dollars that each state has invested in renewable energy from 2002-2017. To create the graph above, the states were ranked based on the amount they have invested in renewable energy and separated into four groups based on their ranking. The mean amount invested in renewable energy was calculated for each of the four groups and graphed against the real investment of each state. The take-away from this graph is that the conditional mean of investment in renewable energy by state is a mediocre predictor of percent change in CO2 emissions and certainly has room for improvement. The rmse was 79.13, meaning that using the conditional mean of investment in renewable energy to predict percent change in CO2 emissions is accurate within 79.13% above or below the actual observation of percent change in CO2 emissions. To put this number in perspective, the total variation of percent change CO2 emissions among the states is 375 percent, meaning that an rmse of 79 is .21 of the data. While there may be a connection between the amount that a state invests in renewable energy sources and the reduction of CO2 emissions over time, the connection is certainly not pronounced. The graph and predictions were made without the outliers of LA or AR. With the outliers of LA and AR, the rmse of the conditional mean was 1553.11, meaning that it is a totally inaccurate predictor of CO2 percent change in emissions.

# How does the presence of a law capping CO2 emissions affect actual CO2 emissions?

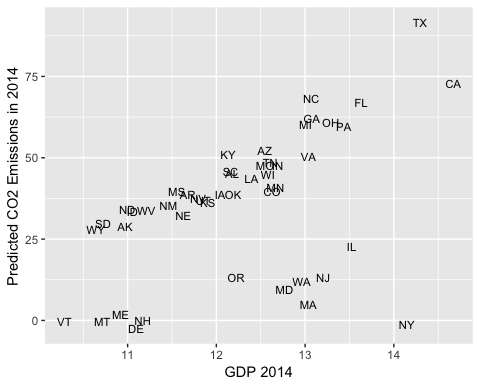
 In comparing states with emission cap laws versus those without, it was found that states with a cap law averaged a 3.26 percent decrease of CO2 emissions from 1980 to 2014 while states without a cap law averaged a 41.95 percent increase in CO2 emissions during that time. While this shows a marked difference between states with and without an emissions cap law, the rmse was 78.61, meaning that the accuracy of this prediction is accurate with 78 percentage points of the actual percent change in emissions. For perspective, this is also about .21 of the total variation of percent change emissions by state, which is 375 percent. Using conditional mean of emission cap law is only slightly more accurate than using the conditional mean of renewable investment as a predictor. It should also be noted that 14 states have an emission cap law and 31 do not, meaning that part of the unreliability could be a consequence of the uneven distribution between the binary variable. In other words, there can be increased variability of percent change emissions of states with an emissions law because there are only 14 data points.

When including the outliers of LA and AR, the mean percent change of states with an emission cap is -3.26 percent, and the mean percent change of states without an emission cap, which includes LA and AR, is 400.73 percent increase in CO2 emissions during the time period. With this dataset, the rmse of the conditional mean is 1604.49, meaning that the prediction is highly inaccurate.

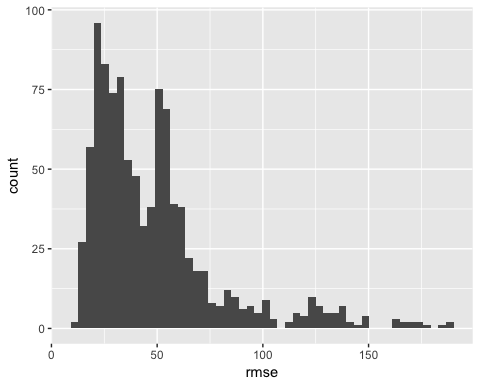
## Predictive Model of CO2 Emissions

## Accuracy of the model

A regression model predicting the emissions of CO2 by state in 2014 is much more accurate than the conditional means, based on its rmse of 26.62. However, it's important to note that this model used actual emissions, in million metric tons of CO2, rather than using percent change in CO2, because it increased the accuracy of the model. This means that the rmse should be interpretted differently; the 26.62 rmse means that the model is accurate in predicting emissions in 2014 within 26.62 million metric tons above or below actual emissions. Total variance between states in emissions in 2014 is 149.8, meaning that the rmse of 26.62 is .18 of total variance, which is only slightly more accurate than the .21 proportion of rmse:variance in the conditional means. Additionally, because the model uses emissions in 2014 instead of percent change of emissions, the regression included LA and AR; while their percent change from 1980-2014 was dramatically higher than the other states, their total emissions in 2014 are not outliers. The regression included whether or not states had an emission cap law, the amount spent on renewable energy, the total amount of unlinked rides using public transportation in 2013, the population in 2014, and the GDP in 2014. Population and GDP were included as controls because they are factors known to influence CO2 emission rates. The emission cap law is statistically significant, based on the p-value of .0033; the other factors are not statistically significant. Emission caps have a negative correlation to emissions, meaning states with an emission cap are predicted to have lower emissions in 2014. The r-squared value is .39, meaning that this model accounts for 39% of the variance within CO2 emissions among states.

 The above graph depicts CO2 emissions predicted from the regression model in terms of GDP from 2014. Each data point is shown as the abbreviation of the correlated state. The relationship shows that GDP and predicted CO2 emissions are positively correlated, meaning that states with higher GDP generally have higher CO2 emissions, a known relationship. There also appears to be two distinguishable lines with the same positive relationship between GDP and predicted emissions; the group of 12 states that is noticeably below the other trend of upward-sloping states is the states that have emission cap laws. In other words, the states with an emission cap law show the same trend of increased GDP increasing predicted emissions, but their total emissions are at lower levels than states without an emission cap law. This is with the exception of Florida and California, both of which have emission cap laws and are in the upper group of states.

# Cross Validation of models



The cross-validation of models, when ran 1000 times at 20% testing data, gives an average rmse of 45.69 and median rmse of 37.55, meaning the distribution is right skewed. As expected, the cross-validated model actually has a higher rmse than the original regression model because it split the data into testing and training, meaning it had less data to work with when informing its predictions. However, the cross-validated model is still more accurate because it is more apt at predicting future data, while the regression model is built to predict the data already present in the dataset. In the figure above, each rmse of the model is plotted, showing that the rmse was generally fairly low, depending on which random observations were categorized into the testing and training datasets each time.

## Case Study: Texas

In 2014, Texas had the highest CO2 emissions: 149.8 million metric tons. To put the various potential interventions to the test, the data was changed separately to show Texas with an emission cap law, with a 20% increase in renewable energy investment, and with a 20% increase in annual public rides to compare these three scenarios with each other and with the actual data. When the regression is run on the actual data, in which TX does not have an emission cap law, has invested $311,337,270 total in renewable energy since 2002, and had 290,399 rides on public transportation in 2013, the regression predicted that TX would emit 91.55 million metric tons.

When TX theoretically had an emission cap law, the predicted CO2 emissions from the regression are 74.34 million metric tons. In other words, the regression model predicts that Texas implementing an emission cap law would reduce annual CO2 emissions by 17.21 million metric tons, or almost 19% of their current emissions. When TX theoretically increased its total investment in renewable energy by 20%, which would be a total investment of 373,604,724, the predicted CO2 emissions for that year are 92.95 million metric tons, which is actually more than the actual predicted emissions of Texas. When TX theoretically increased its annual rides on public transportation by 20%, the regression predicted 90.83 million metric tons would be emitted that year, which is a decrease of .72 million metric tons. Clearly, the regression predicts that the most effective way to dramatically reduce CO2 emissions in Texas would be to enact an emission cap law. Other forms of investment are more expensive and only nominally reduce emissions, or may even increase emissions.

# Take-aways

Overall, it seems that emission cap laws have a more significant impact on CO2 emissions by state than investments in renewable energy or the use of public transportation. The models included state GDP and population to control for these variables, since they are known to be highly correlated to total CO2 emission by state. The outliers of Louisiana and Arkansas show the extreme effects that increased industry can have on CO2 emissions in a state, although the fact that the percent change variables are outliers and the total emissions in 2014 of the two states are not outliers implies that these massive changes are actually still comparable to what other states were already annually emitting in CO2. Under the assumption that the goal is to minimize CO2 emissions, government agencies should invest their time and energy into enacting an emission cap law in the state, or improving the rigor of the law if it already has one in place, instead of investing in renewable energy or public transportation. While the emission cap law has a statistically significant negative relationship to CO2 emissions, it is still contested in some states based on the burden it places on corporations and industries. Theoretically, enacting an emission cap law would be free for the state government; however, the true cost would be the loss of industry on the side of the corporations. This loss could be mitigated through new, cleaner technologies and cap and trade techniques that allow CO2-heavy industries to "trade" for their increased emissions. Other CO2 reduction options like increasing government investment in renewable energy or increasing public transportation are more costly and not as strongly correlated to CO2 emission reduction. Therefore, the state legislators should enact or strengthen an emission cap law and allow for cap and trade to ease some of the burden on industries.